SOME ADAPTIVE MONTE CARLO METHODS IN STATISTICAL MECHANICS

J. E. Gubernatis
Theoretical Division
Los Alamos National Laboratory
jeg@kaiser.lanl.gov

OUTLINE

- Some Monte Carlo Concepts
- Some Methods
  - Multi-canonical
  - Tempering
- Some Remarks

SOME MONTE CARLO CONCEPTS

Markov chain Monte Carlo methods are used in statistical mechanics to sample a configuration C with a probability

$$P(C) = \frac{e^{-\beta E(C)}}{Z(\beta)}$$

to replace multivariate integrations

$$\langle f \rangle = \sum_C f(C) P(C)$$

by simple averages

$$\langle f \rangle \approx \bar{f} \equiv \frac{1}{M} \sum_{i=1}^{M} f(C_i) \equiv \frac{1}{M} \sum_i f_i$$

- $P(C)$ is called the canonical (Boltzmann) distribution.
- $Z(\beta)$ is called the partition function.

$$Z(\beta) = \sum_C e^{-\beta E(C)}$$

- β is the inverse of the temperature.
- $E(C)$ is the energy of configuration C.

The Metropolis algorithm which defines

$$P(C' \leftarrow C) = \min[1, P(C')/P(C)]$$

is the standard method used to produce the Markov Chain.

For the Boltzmann distribution, the transition probability reduces to

$$P(C' \leftarrow C) = \min[1, e^{-\beta \Delta E}]$$

where $\Delta E = E(C') - E(C)$. Note the independence from $Z(\beta)$.

In statistical mechanics,

$$E(C) \sim \mathcal{O}(N)$$

where N is the number of particles; however, the changes in the configurations are constructed so that typically

$$\Delta E \sim \mathcal{O}(1)$$

Because the changes are small, correlations exist between successive measurements and the measured variance of a physical quantity

$$\sigma^2_{\rm mc} = \frac{1}{M-1}\sum_{i=1}^{M}(\bar{f} - f_i)^2$$

becomes related to the true variance by

$$\sigma^2_{\rm true} = \tau_{\rm mc}\,\sigma^2_{\rm mc}$$

where $\tau_{\rm mc}$ is an auto-correlation time measured along the chain.

Under typical conditions of interest, $\tau_{\rm mc}$ is often $\sim 10^2 - 10^8$ for reasons as simple as the Monte Carlo simulations simulating physical processes that have long relaxation time.

Example, near a first-order phase transition

$$\tau_{\text{phys}} \propto e^{cE_b}$$
$$E_b \propto L^{D-1}$$

where L is the "length" of the system and D is its dimension.

The same scaling with system size is observed in standard Monte Carlo simulations.

We cannot change τ_{phys} but we can try to change τ_{mc}. Recently, several adaptive Monte Carlo methods for sampling multi-modal distribution, like the bi-modal one observe near a first-order transition, have been proposed.

SOME ADAPTIVE METHODS

There seems to be two classes of methods:

1. Multi-canonical
   a. replace the sampling from multi-modal Boltzmann distribution by the sampling from a smoother (uni-modal) non-Boltzmann distribution.
   b. Examples: umbrella sampling, multi-canonical sampling, entropic sampling, etc.
2. Tempering
   a. treat the temperature as a random variable to facilitate movement over barriers
   b. Examples
      i. expanded ensemble, simulated tempering, adaptive tempering, etc.
      ii. J-walking, exchange Monte Carlo, etc.

The Multi-Canonical Method*

$$\begin{aligned}
\langle f \rangle &= \sum_C f(C) e^{-\beta E(C)} / Z(\beta) \\
&= \sum_C f(C) P(C) \\
&= \sum_E f(E) w(E) e^{-\beta E} / Z(\beta) \\
&= \sum_E f(E) P(E) \\
&= \sum_E f(E) \frac{P(E)}{I(E)} I(E) \\
&\approx \frac{1}{M} \sum_i f_i \frac{P_i}{I_i}
\end{aligned}$$

How does one choose $I(E)$?

* B.A. Berg and T. Nehaus, Phys. Lett. B **267**, 249 (1991); Phys. Rev. Lett. **68**, 9 (1992).

The Algorithm.

Require

$$I(E) = \frac{w(E)e^{-H[E]}}{Z'}$$

where

$$Z' = \sum_E w(E)e^{-H[E]}$$

Note that $I(E)$ is independent of β!

Observe that if

$$w(E)e^{-H[E]} = 1$$

or equivalently

$$H[E] = \ln w(E)$$

then

$$I(E) = \frac{1}{Z'} = \frac{1}{N_E}$$

Estimate $I(E)$ from a simulation

$$I(E) = \frac{w(E)e^{-H[E]}}{Z'} \approx h(E)$$

or equivalently

$$\ln w(E) = H[E] + \ln Z' h(E)$$

Start with $H[E] = \beta E$,

1. Metropolis sample with $I(E)$ to estimate $h(E)$
2. $H[E] \leftarrow H[E] + \ln N_E h(E)$
3. Repeat

Expanded Ensemble*

What we have been calling $P(C)$ or $P(E)$ are really conditional probabilities.

$$P(C) \to P(C|\beta)$$

Consider a joint probability

$$P(C,\beta) = P(C|\beta)P(\beta)$$

where $P(\beta)$ is to be determined. It follows that

$$P(C) = \sum_{\beta} P(C,\beta) = \sum_{\beta} P(C|\beta)P(\beta)$$

$$\begin{aligned}\langle f \rangle &= \sum_{C} f(C)P(C|\beta) \\ &= \sum_{C} \frac{P(C|\beta)}{I(C)} I(C)\end{aligned}$$

Take

$$I(C) = \sum_{\beta} P(C|\beta)P(\beta)$$

where we note that $I(C)$ is independent of β.

How does one choose $P(\beta)$?

---

* A.P. Lubartsev et al., J. Chem. Phys. **96**, 1776 (1992); E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).

The Algorithm.

Choose a finite number of values of β designated by β_i.

Require

$$P(C, \beta_i) = \frac{e^{-\beta_i E} e^{-\eta_i}}{Z}$$

which implies

$$Z = \sum_i e^{-\eta_i} \sum_C e^{-\beta_i E} = \sum_i Z(\beta_i) e^{-\eta_i}$$

$$P(\beta_i) = \sum_C P(C, \beta_i) = \frac{Z(\beta_i) e^{-\eta_i}}{Z}$$

Observe that if

$$Z(\beta_i) e^{-\eta_i} = 1$$

then

$$P(\beta_i) = \frac{1}{Z} = \frac{1}{N_\beta}$$

Estimate $P(\beta)$ from a simulation

$$P(\beta_i) = \frac{Z(\beta_i)e^{-\eta_i}}{Z} \approx h(\beta_i)$$

or equivalently

$$Z(\beta_i) \approx \eta_i + \ln Z h(\beta_i)$$

Start with $\eta_i = -\ln Z(\beta_i)$ estimated by some means,

1. Perform a Monte Carlo simulation
   a. Select a β_i with probability $P(\beta)$
   b. Metropolis sample with $P(C|\beta_i)$
2. $\eta_i \leftarrow \eta_i + \ln N_\beta h(\beta_i)$
3. Repeat

Exchange Monte Carlo*

$$P(C|\beta) \to P(C_1, C_2, \ldots, C_{N_\beta} : \beta_1, \beta_2, \ldots, \beta_{N_\beta})$$
$$\to \prod_i P(C_i|\beta_i)$$

The algorithm

For each β_i, simultaneously do a standard Monte Carlo sampling.

Periodically, exchange a pair of configurations i and j with probability

$$\min[1, \exp(-\Delta\beta\Delta E)]$$

where

$$\Delta\beta = \beta_i - \beta_j$$
$$\Delta E = E(C_j) - E(C_i)$$

The algorithm as used to date is not adaptive but can be made so by attempting to adjust the probability with which temperatures pairs are chosen.

---

* K. Hukushima and K. Nemoto, unpublished.

REMARKS

• The various adaptive sampling methods make orders of magnitude reduction in the auto-correlations times.

⊙ The fewer the modes the better.

• They also allow estimation of the partition function $Z(\beta)$, the free-energy $-\ln Z(\beta)/\beta$, and the entropy $\ln w(E)$.

⊙ Usually a simulation at one value of β determines results at many values of β.

• The methods are *ad hoc*.

⊙ The recursive equation can be interpreted in the context of Bayesian learning methods.

$$P(\theta|D_1, D_2, \ldots, D_{n+1}) \propto P(\theta|D_1, D_2, \ldots, D_n)P(D_{n+1}|\theta)$$

• One computational bottleneck is the density estimation step.

  ⊙ This is the subject of research by R. Silver and myself.

• Another bottleneck is the continued use of standard site-by-site Monte Carlo simulation methods.

• In actuality, the multi-canonical method and expanded ensemble method are special cases of a more general method.

$$e^{-H[E_i]} \rightarrow e^{-\beta_i E_i}$$

• Although more general, it is unclear if these adaptive methods are better in specific applications than other recently established methods, for example, cluster algorithms, hybrid Monte Carlo, Gibbs ensemble, etc.